

Original Communication

Systematic evaluation of sensitivity and specificity of sibship determination by using 15 STR loci

Chang En Pu MS (Scientist Specialist and Section Chief)^{a,*},
Adrian Linacre PhD (Senior Lecturer)^b

^a Forensic DNA Section, Scientific and Technical Research Center, Ministry Justice Investigation Bureau,
No. 74 Chun_hya Road, Hsin-tien, Taipei County 231, Taiwan

^b Centre for Forensic Science, Department of Pure and Applied Chemistry, University of Strathclyde, Glasgow, UK

Received 17 June 2007; received in revised form 30 August 2007; accepted 31 December 2007

Available online 11 April 2008

Abstract

Paternity disputes and other forms of kinship testing are routinely resolved using short tandem repeat (STR) DNA loci. Sibship determination is encountered in instances where the DNA profiles of two individuals are compared to determine if they are siblings. If either parent is available for testing then the situation is simplified but if neither parent of the two individuals is available for DNA testing, a combined sibling indices (CSI) for the determination of sibship between two people can be determined. Support for kinship is also based upon the sharing of alleles, particularly when both alleles are shared at the same locus, termed two-allele-sharing-loci (TASL). We report on the combination of CSI and TASL to enhance the determination of sibship. The 15 STR loci that comprise the Identifiler[®] loci were applied to three populations using pairs of full siblings or unrelated pairs. Based upon the data obtained, the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) can be applied to determining whether two DNA profiles come from full or non-sibling pairs. This report highlights the problems inherent in this form of kinship testing and recommends a combination use of CSI and TASL for sibship determination.

© 2008 Elsevier Ltd and FFLM. All rights reserved.

Keywords: Short tandem repeat (STR); Kinship; Sibship; Combined sibship index (CSI); Sensitivity; Specificity

1. Introduction

Sibship determination is encountered in instances such as linking human remains to a relative and when neither of the biological parents of the two individuals is available for testing. Theoretically at any one locus there is a 0.5 chance that two siblings will share one allele, a 0.25 chance that they will share neither allele and a 0.25 chance that they will share both alleles. The chance that two unrelated individuals share either one or both alleles at any one locus is dependent upon the frequency of the alleles.^{1,2} A probability of sibship can be determined based upon the frequency of the matching alleles

in the population and will increase when a high number of high discriminating loci are examined.³ Uncertainty using DNA testing to resolve sibship increases if the parents were heterozygous rather than homozygous.^{1,4} There is no evidence of sibship if there is no two-allele-sharing-locus (TASL) between the two profiles,⁵ however confidence is increased if a number of TASL existed.

Based upon the degree of sharing of alleles between two DNA profiles it is possible to determine a combined sibship indices (CSI).¹ When the index is less than 1 the two individuals might not be related as siblings. If the index is over 1 then the data supports the existence of a sibling relationship. Other cut-off point have been recommended such a $CSI \geq 3$.² These figures are guides as it was found that 1.6% of random pairs of DNA profiles had CSI greater than 1² when using 15 STR loci. In a study using 16 STR

* Corresponding author. Tel.: +886 2 29112241x3740; fax: +886 2 291 38599.

E-mail address: pu_macros@yahoo.com.tw (C.E. Pu).

loci 0.1% of unrelated pairs of DNA profiles were found to have CSI > 100 and 0.01% for CSI > 1000.⁶ Using the 15 STR loci used in the Identifiler® kit, none of the non-sibling pairs were found with CSI ≥ 1, while all sibling pairs have CSI > 10.⁷ In a different study using the same loci, 6.06% of sibling pairs exhibited CSI < 1 and 9.1% of random unrelated pairs had CSI > 1.⁸ The variation of percentage of random pairs with CSI > 1 found above maybe owing to the different levels of consanguinity that might be present in those populations.

We have extended the studies using the 15 STR comprising the Identifiler® loci to study three populations using 357,630 full sibling pairs and 178,815 non-sibling pairs generated from DNA profiles of random population. Using this high number of sample pairs, it is possible to evaluate the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of the 15 STR loci for discriminating between full and non-siblings. The relationship between the CSI values and TASL is reported.

2. Materials and methods

DNA profiles from a Chinese population were obtained using the ABI AmpFISTR Identifiler® PCR Amplification Kit (Applied Biosystem, Foster City, CA, USA). The STR products were analyzed with an ABI Prism 3100 XL Genetic Analyzer.

STR profiles (450) of random members of the Taiwanese Chinese population were processed by Microsoft Excel Macros controlled by Visual Basic program written by authors of this study. Every member of the population was paired with every other member to form random pairs, e.g. for Chinese population in this study (450 × 449)/2, equaling 101,025 pairs, were made. Every pair was set to have two children, resulting in 202,050 sibling pairs being generated. DNA profiles from the 15 STR loci from a Caucasian (n = 301) and American African (n = 256) popula-

tion were obtained from short tandem repeat DNA internet database.⁹ These data were processed in the same way as those of the Taiwanese population to generate sibling pairs and random pairs. Inevitably for computer based populations no account of substructure is made and mating occurs randomly. The combined sibship indices (CSI) were calculated for each simulated sibling pairs or random pairs by using standard formulae,¹ and allele frequency tables used for calculation of CSI were adjusted by using 5/2N rule.¹⁰

The rate of false negatives equaled the percentage of real sibship testing cases (in this study the simulated sibling pairs) that would be excluded based upon any given cut-off point of CSI or TASL. The rate of false positives equaled the percentage of random pairs of DNA profiles where their CSI or TASL was greater than any recommended cut-off value. The sensitivity of the test is based upon 1 – the % of false negatives, the specificity of the test is based upon 1 – the % of false positives, the positive predictive value (PPV) = the proportion of subjects correctly identified as siblings and the negative predictive value

Table 1
Maximum and minimum CSI for three populations using 15 STR systems

Population	Profiles	Type	Pairs	Maximum	Minimum
Chinese	450	Simulated siblings	202,050	4.06E+16	1.68E–05
		Random pairs	101,025	4.79E+04	2.33E–09
Caucasians	301	Simulated siblings	90,300	1.23E+14	3.42E–04
		Random pairs	45,150	1.52E+04	6.38E–09
African Americans	256	Simulated siblings	65,280	3.54E+14	9.52E–04
		Random pairs	32,640	6.24E+08	3.43E–09

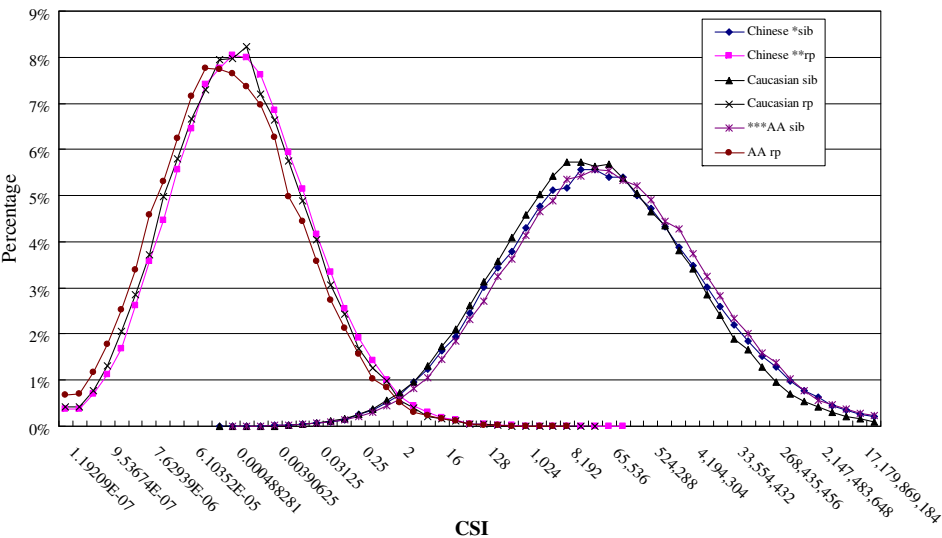


Fig. 1. Ratio distribution of CSI for three populations (* sib: siblings; ** rp: random pairs; *** AA: African Americans).

Table 2
Evaluation of sensitivity, specificity, PPV, NPV and optimal CSI cut-off points for sibship determination using 15 STR system

Population	Chinese						Caucasians						African Americans					
CSI	^a SEN (%)	^b SPE (%)	^c Abs- <i>X</i>	^d Sqrt- <i>Y</i>	PPV (%)	NPV (%)	SEN (%)	SPE (%)	Abs- <i>X</i>	Sqrt- <i>Y</i>	PPV (%)	NPV (%)	SEN (%)	SPE (%)	Abs- <i>X</i>	Sqrt- <i>Y</i>	PPV (%)	NPV (%)
0.03125	99.849	87.905	0.1194	1.3303	89.195	99.828	99.868	88.977	0.1089	1.3376	90.059	99.852	99.859	90.291	0.0957	1.3463	91.139	99.844
0.067	99.747	91.538	0.0821	1.3538	92.180	99.724	99.750	92.330	0.0742	1.3592	92.860	99.730	99.755	93.269	0.0649	1.3657	93.679	99.738
0.125	99.615	93.795	0.0582	1.3682	94.136	99.592	99.590	94.454	0.0514	1.3726	94.725	99.568	99.619	95.150	0.0447	1.3776	95.358	99.601
0.25	99.364	95.713	0.0365	1.3796	95.864	99.339	99.338	96.146	0.0319	1.3825	96.265	99.316	99.403	96.713	0.0269	1.3869	96.799	99.386
0.5	99.012	97.131	0.0188	1.3870	97.184	98.993	98.952	97.420	0.0153	1.3886	97.459	98.936	99.096	97.736	0.0136	1.3918	97.766	99.084
1	98.500	98.125	0.0037	1.3904	98.132	98.494	98.392	98.410	0.0002	1.3916	98.409	98.392	98.660	98.585	0.0008	1.3947	98.586	98.659
2	97.816	98.788	0.0097	1.3902	98.777	97.837	97.659	98.966	0.0131	1.3904	98.952	97.689	98.068	99.093	0.0102	1.3942	99.084	98.088
3	97.298	99.050	0.0175	1.3884	99.033	97.345	97.174	99.207	0.0203	1.3887	99.191	97.230	97.603	99.277	0.0167	1.3922	99.265	97.642
10	95.151	99.613	0.0446	1.3776	99.595	95.358	94.876	99.670	0.0479	1.3761	99.653	95.110	95.760	99.691	0.0393	1.3823	99.678	95.920
10.3	95.074	99.624	0.0455	1.3771	99.606	95.288	94.791	99.672	0.0488	1.3755	99.655	95.033	95.695	99.694	0.0400	1.3819	99.681	95.861
33	91.953	99.852	0.0790	1.3574	99.839	92.542	91.466	99.880	0.0841	1.3543	99.869	92.129	92.785	99.893	0.0711	1.3634	99.885	93.264
100	87.735	99.935	0.1220	1.3298	99.926	89.068	87.009	99.949	0.1294	1.3252	99.941	88.497	88.919	99.942	0.1102	1.3377	99.935	90.019
150	85.825	99.953	0.1413	1.3174	99.946	87.580	85.012	99.969	0.1496	1.3123	99.964	86.962	87.109	99.966	0.1286	1.3259	99.961	88.578
200	84.429	99.958	0.1553	1.3084	99.951	86.522	83.627	99.978	0.1635	1.3034	99.974	85.928	85.846	99.972	0.1413	1.3177	99.968	87.598
300	82.331	99.969	0.1764	1.2951	99.963	84.980	81.349	99.980	0.1863	1.2889	99.976	84.278	83.791	99.975	0.1618	1.3045	99.971	86.049
330	81.818	99.971	0.1815	1.2918	99.965	84.611	80.788	99.984	0.1920	1.2854	99.981	83.882	83.346	99.975	0.1663	1.3016	99.971	85.720
500	79.492	99.976	0.2048	1.2773	99.970	82.979	78.303	99.987	0.2168	1.2700	99.983	82.170	81.150	99.975	0.1883	1.2877	99.970	84.137
1000	75.218	99.984	0.2477	1.2512	99.979	80.137	73.754	99.989	0.2623	1.2425	99.985	79.209	77.004	99.982	0.2298	1.2620	99.976	81.300
Minimum	75.218	87.905	0.0037	–	89.195	80.137	73.754	88.977	0.0002	–	90.059	79.209	77.004	90.291	0.0008	–	91.139	81.300
Maximum	99.849	99.984	–	1.3904	99.979	99.828	99.868	99.989	–	1.3916	99.985	99.852	99.859	99.982	–	1.3947	99.976	99.844

^a SEN: sensitivity.

^b SPE: specificity.

^c Abs-*X*: Abs(SEN – SPE).

^d Sqrt-*Y*: Sqrt(SEN² + SPE²).

(NPV) = the proportion of subjects correctly identified as non-siblings.⁴

The optimum CSI cut-off points were obtained by minimizing, using sensitivity – specificity,¹¹ and maximizing by using the square root of (sensitivity² + specificity²).¹²

3. Results and discussion

3.1 Ratio distribution of CSI for three populations

The CSI ratio distribution of simulated sibling pairs and random pairs are depicted in Fig. 1. Bipolar models with a widespread ratio distribution were found for all of the three populations. From the data obtained, 1.500%, 1.608% and 1.340% of CSI values from simulated sibling pairs were found to be less than 1 for Chinese, Caucasians and African Americans populations, respectively, and 1.875%, 1.590% and 1.415% of random pairs were found to have CSI larger than 1 (Table 2). Simulated siblings pairs with very low CSI values (1.68E – 05 for Chinese population) and random pairs with very high CSI values (6.24E + 08 for African American population) were observed in this

study. Other extreme CSI values were also found in Table 1. While it is never possible to be definitive on sibship, a large uncertainty value can exist. It was proposed in⁸ that when two DNA profiles produce CSI values between 0.67 and 10 further loci should be analyzed. In this present study the lowest CSI value for a sibling pair was significantly less than 0.067 and CSI for random pairs was much larger than 10.3. Our data indicates that simply applying a minimum CSI requirement may result in false exclusions if set too high or false inclusions if too low.

3.2 Sensitivity, specificity, PPV and NPV under different CSI cut-offs

The ability of the DNA test to correctly classify kinship testing results into two categories (sibling or not sibling) is assessed by specificity and sensitivity.¹³ In Table 2 the sensitivity and specificity using a range of CSI values is illustrated. CSI cut-off values at 0.067, 3 and 10.3 are used as they follow previous recommendations.^{2,4,8,14} As the CSI cut-off values increased there was a corresponding decrease in sensitivity and increase in specificity. When adopting a

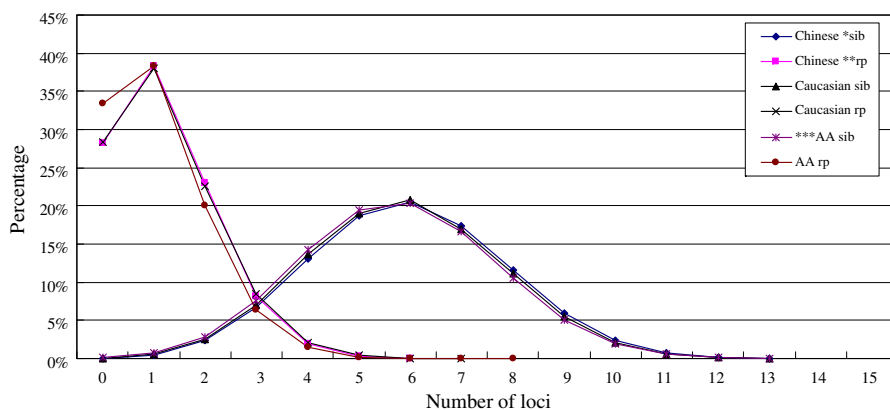


Fig. 2. Ratio distribution of TASL for three populations (* sib: siblings; ** rp: random pairs; *** AA: African Americans).

Table 3
Sensitivity, specificity, PPV and NPV versus two allele sharing cut-off points

Population	Chinese				Caucasians				African Americans			
Loci	SEN (%)	SPE (%)	PPV (%)	NPV (%)	SEN (%)	SPE (%)	PPV (%)	NPV (%)	SEN (%)	SPE (%)	PPV (%)	NPV (%)
1	99.952	28.147	58.178	99.831	99.956	28.430	58.274	99.837	99.917	33.447	60.021	99.753
2	99.445	66.459	74.778	99.171	99.419	66.412	74.747	99.129	99.226	71.752	77.841	98.933
3	97.075	89.484	90.226	96.835	96.904	88.913	89.733	96.632	96.437	91.801	92.165	96.264
4	90.414	97.689	97.507	91.064	89.889	97.415	97.205	90.595	88.932	98.254	98.074	89.876
5	77.376	99.662	99.566	81.499	76.230	99.526	99.382	80.720	74.710	99.752	99.669	79.775
6	58.641	99.957	99.927	70.733	57.224	99.940	99.896	70.026	55.277	99.963	99.934	69.090
7	38.132	99.997	99.992	61.778	36.412	99.996	99.988	61.127	34.910	99.994	99.982	60.571
8	20.828	100.000	100.000	55.812	19.542	100.000	100.000	55.414	18.284	99.997	99.983	55.030
9	9.295	100.000	100.000	52.437	8.475	100.000	100.000	52.212	7.736	100.000	100.000	52.012
10	3.284	100.000	100.000	50.835	2.907	100.000	100.000	50.737	2.627	100.000	100.000	50.666
11	0.879	100.000	100.000	50.221	0.788	100.000	100.000	50.197	0.700	100.000	100.000	50.176
12	0.169	100.000	100.000	50.042	0.155	100.000	100.000	50.038	0.146	100.000	100.000	50.036
13	0.025	100.000	100.000	50.006	0.025	100.000	100.000	50.006	0.023	100.000	100.000	50.006
14	0.003	100.000	100.000	50.001	0.002	100.000	100.000	50.000	0.000	100.000	100.000	50.000
15	0.000	100.000	100.000	50.000	0.000	100.000	100.000	50.000	0.000	100.000	100.000	50.000

C.E. Pu, A. Lincare / *Journal of Forensic and Legal Medicine* 15 (2008) 329–334

333

simple CSI value of 1, below which indicates non-sibling and above supports a sibling pair, then for the three populations the sensitivity was 98.500% and the specificity was 98.125% for the Chinese population, for Caucasian population the sensitivity was 98.392% and the specificity was 98.410% and for African American population the sensitivity was 98.660% and the specificity was 98.585%. The PPV and NPV were all over 98% for the three populations.

3.3 Ratio distribution, sensitivity and specificity of TASL

The TASL ratio of simulated sibling pairs and random pairs is depicted in Fig. 2. Bipolar models were found for all of the three populations.

The sensitivity and specificity for 1–15 of TASL are illustrated in Table 3. As the number of cut-off loci increased there was a corresponding decrease in sensitivity and increase in specificity. At the TASL cut-off ≥ 5 , the specificity was greater than 99% for all of the populations, but the sensitivity was only around 75%. Based upon these data, TASL alone is not an optimal screening method for the sibship determination. A previous instance was reported where no TASL were found for a pair of DNA profiles⁵ leading to the exclusion of sibship; based upon our data (Table 3) the NPV was 99.837% (for Caucasians) regardless of CSI value. The use of a TASL value of 1 as the cut-off only in place of CSI calculating was found to be more sensitive and easy for sibship screen of mass comparing, since the sensitivity was 99.95% compared to a sensitivity value of 98.500% for a CSI of 1 for Chinese, and the comparing could be done without the help of computer. When pairs were screened then CSI calculation applied afterward and Table 4 could be used to determine the specificity.

3.4 Evaluation of synergy effects of two criteria for the sibship determination

The sensitivity and specificity of both the CSI and TASL values were combined. Values of CSI varying from 0.125 to 100 were analyzed against the values of 1–9 for TASL (Table 4). A combination of these two data sets increased the confidence of in an exclusion or inclusion. A CSI of 0.125 and TASL of five resulted in a specificity of 99.727% for Chinese population, 99.626% for the Caucasian population and 99.804% for African American population. Instances of medium CSI values were reflected as the true specificity, e.g. when CSI = 3 and TASL = 5, resulted in the specificity increasing to 99.885% instead of 99.050% (see Table 2, when based on CSI = 3 only), 99.825% instead of 99.207% for Caucasian population and 99.905% instead of 99.277% for African American population.

The data obtained in this study indicates that values of CSI and TASL to be adopted as the cut-offs may vary from population to population. The values used may also vary if the DNA test is used for identification of human remains or in criminal cases if the burden of proof required also

varies. The ideal situation to resolve sibship cases with greater confidence would be to use more autosomal STR loci if possible or to use mitochondrial DNA analysis or STR loci on either of the sex determining chromosomes.

4. Conclusion

The 15 loci STR core set system used in this study could not distinguish all 357,630 full sibling pairs from 178,815 non-sibling pairs, as a small percentage of false positives and false negatives were found. The use of CSI and TASL together increased the confidence in sibship determination. Using TASL = 1 for the first-step-screening for disaster cases was also proposed.

Acknowledgments

The authors wish to thank Wu F.C., Wu G.C., Cheng M.Y., Meng L.M. and Cheng S.M. for their kindly help for genotyping all the local samples.

References

- Wenk RE, Traver M, Chiafari FA. Determination of sibship in any two persons. *Transfusion* 1996;**36**:259–62.
- Tzeng CH, Lyuu JY, Chen YR, Hu HY, Lin JS, Wang SY, et al. Determination of sibship by PCR-amplified short tandem repeat analysis in Taiwan. *Transfusion* 2000;**40**:840–5.
- Wenk RE, Chiafari FA. Distinguishing full siblings from half-siblings in limited pedigrees. *Transfusion* 2000;**40**:1148–9.
- Gaytmenn R, Hildebrand DP, Sweet D, Pretty IA. Determination of the sensitivity and specificity of sibship calculations using AmpFISTR profiler plus. *Int J Legal Med* 2002;**116**:161–4.
- Henke J, Henke L. Which short tandem repeat polymorphisms are required for identification? Lessons from complicated kinship cases. *Croat Med J* 2005;**46**(4):593–7.
- Thomson JA, Ayres KL, Pilotti V, Barrett MN, Walker JIH, Debenham PG. Analysis of disputed single-parent/child and sibling relationships using 16 STR loci. *Int J Legal Med* 2000;**115**:128–34.
- Reid TM, Wolf CA, Kraemer CM, Lee SC, Baird ML, Lee RF. Specificity of sibship determination using the ABI identifier multiplex system. *J Forensic Sci* 2004;**49**:1–3.
- Giroti RI, Verma S, Singh K, Malik R, Talwar I. A gray zone approach for evaluation of 15 short tandem repeat loci in sibship analysis: a pilot study in Indian subjects. *J Clin Forensic Med*. doi:10.1016/j.jcfm.2006.08.002.
- Short tandem repeat DNA internet database. <http://www.cstl.nist.gov/div831/strbase>.
- Butler JM. Kinship and parentage testing. In: Butler JM, editor. *Forensic DNA typing*. MA, USA: Elsevier Academic Press; 2005. p. 529–37.
- Find 'best' cut-offs for maximizing sensitivity/specificity. Available from: <http://www.natureserve.org/visitLocal/conftraining/edworkshop/rochelp.pdf>. 2007.
- Zou KH, Wells III WM, Kikinis R, Warfield SK. Three validation metrics for automated probabilistic image segmentation of brain tumors. *Statist Med* 2004;**23**:1259–82.
- Kirkwood BR, Sterne JAC. Measurement error: assessment and implications. In: Kirkwood BR, Sterne JAC, editors. *Medical statistics*. MA, USA: Blackwell Publishing; 2003. p. 429–46.
- AABB, 2002 parentage testing annual report. MD, USA: American Association of Blood Banks; 2007. http://www.aabb.org/Documents/Accreditation/Parentage_Testing_Accreditation_Program/ptannrpt02.pdf.